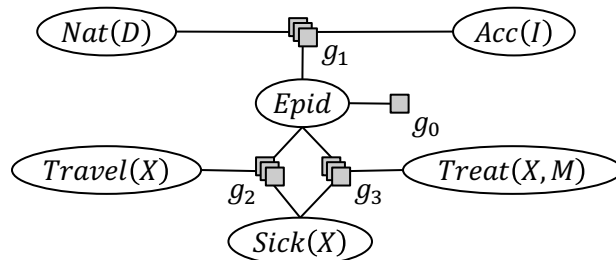


Lifted Inference

Probabilistic Relational Models And Differential Privacy

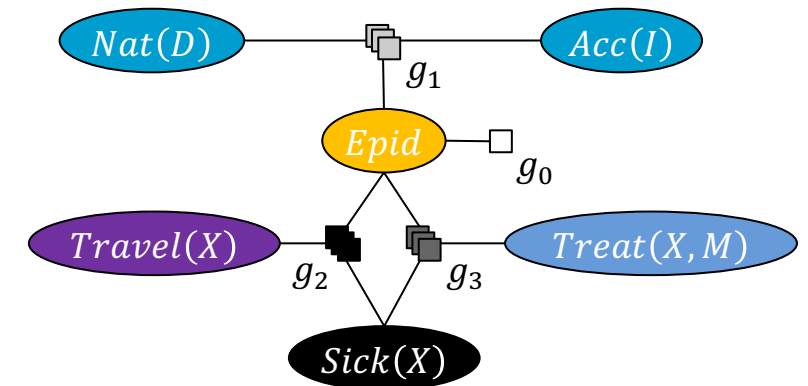


10 $Presents(X, P, C) \Rightarrow Attends(X, C)$

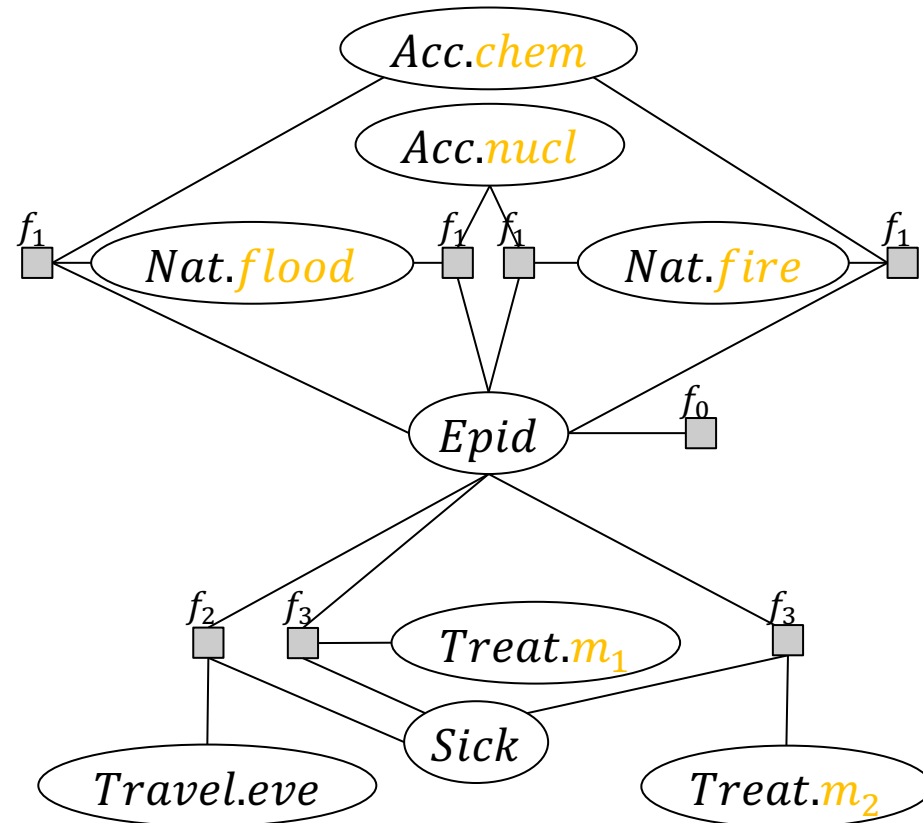
3.75 $Publishes(X, C) \wedge FarAway(C) \Rightarrow Attends(X, C)$

Probabilistic Relational Models

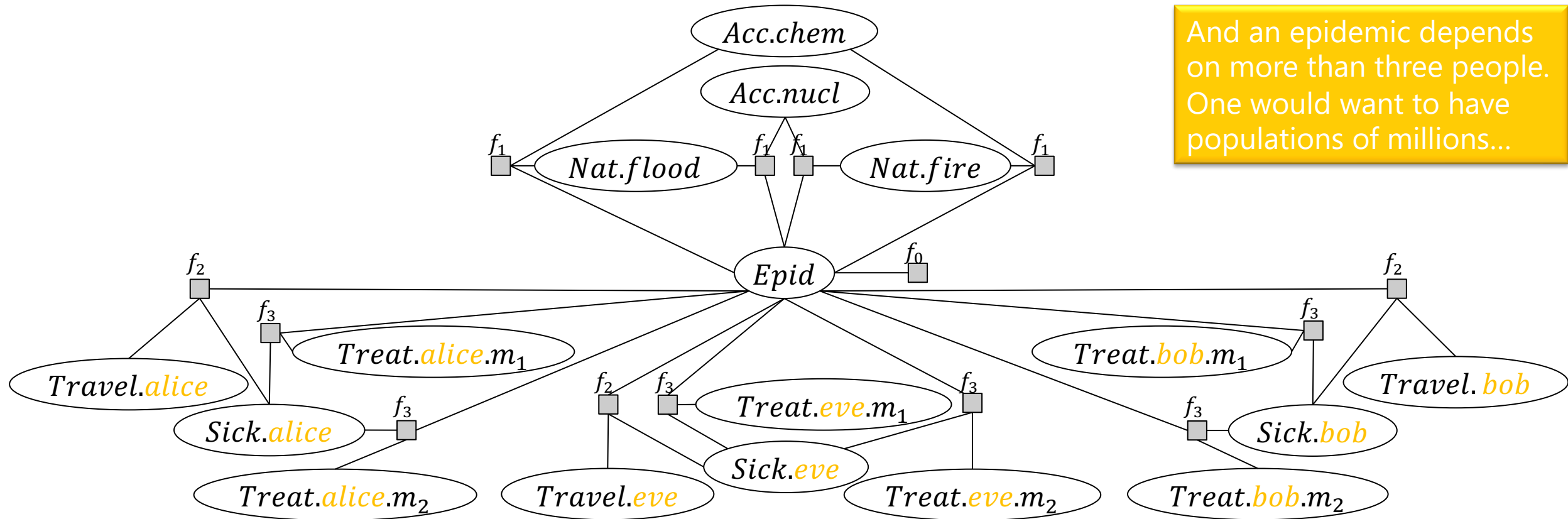
- Accounting for
 - Uncertainty
 - Relations of objects
- Introduces logical variables and constants to PGMs
- Efficient reasoning over indistinguishable objects
- Generative model
- How to obtain a (DP) PRM?
- How to perform DP inference?
 - Do out-of-the-box lifted inference algorithms suffice for DP?
- How to construct DP data sets?



Problem: Adding Objects and Relations in Propositional Formalisms Clunky...



Problem: ... and Models Explode with Them

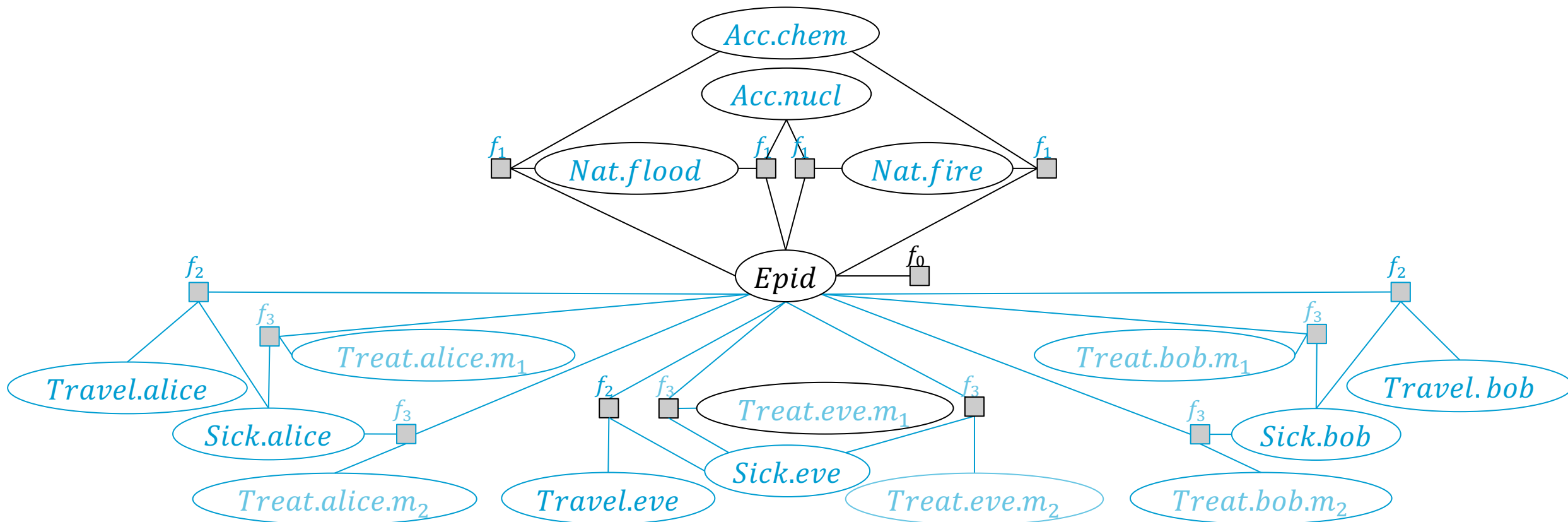


And an epidemic depends on more than three people. One would want to have populations of millions...

$13 \cdot 2^3 + 2 = 104$ entries in 14 factors, 17 variables

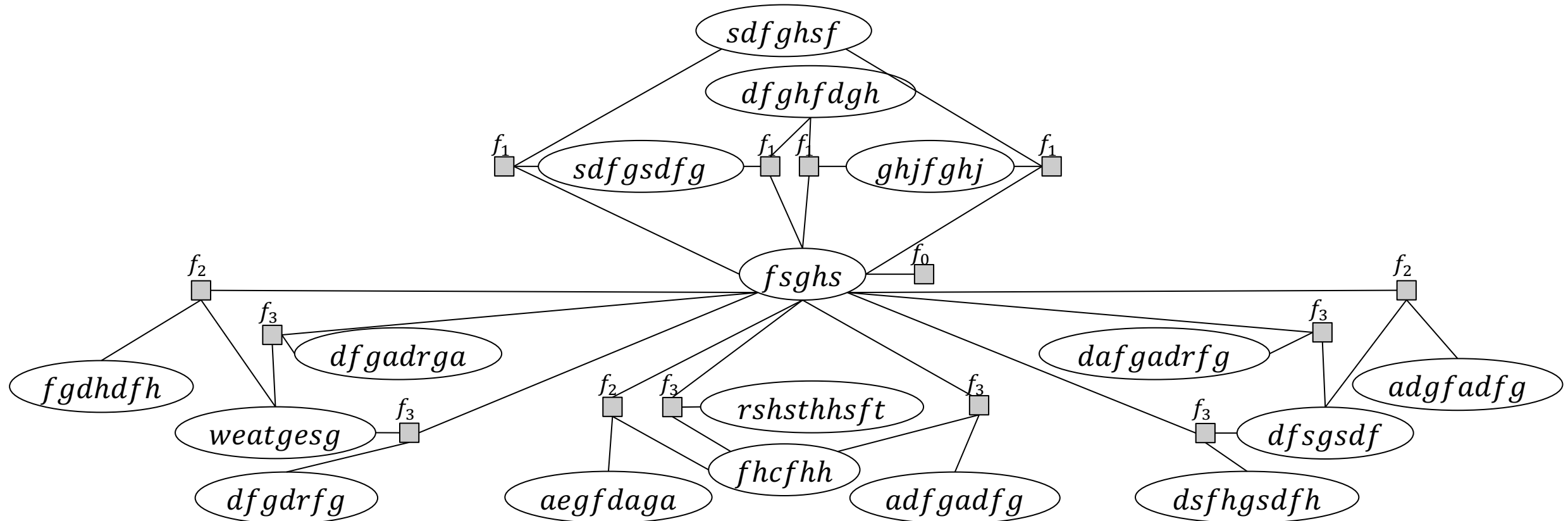


Propositional \rightarrow First-order View



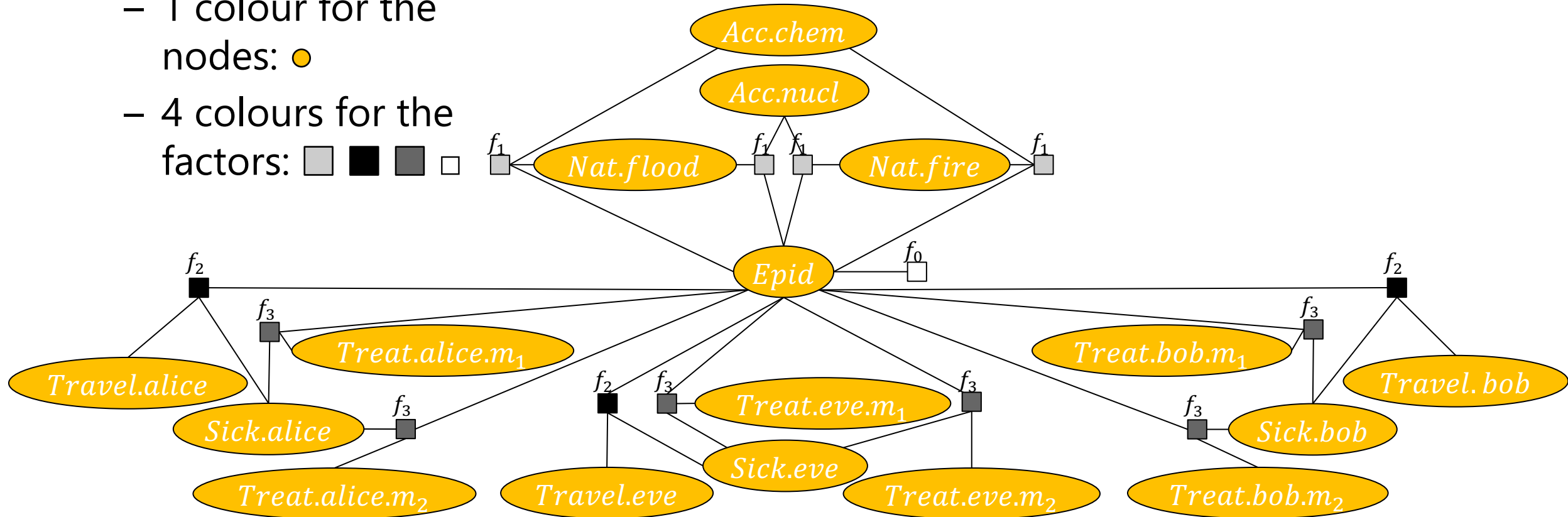
Symmetries in graph / relations in scenario

Problem: ... learned from Data more like this



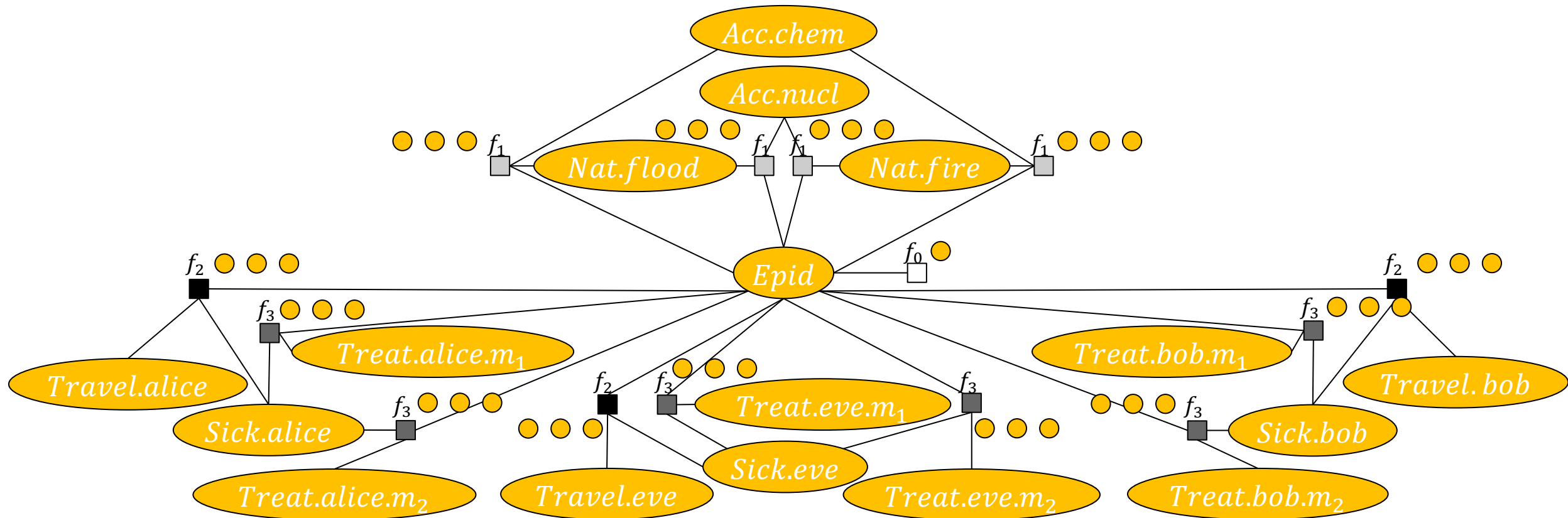
Compression

- Initialisation: Colour nodes and factors
 - 1 colour for the nodes: ●
 - 4 colours for the factors: ◻ ◼ ◾ ◿



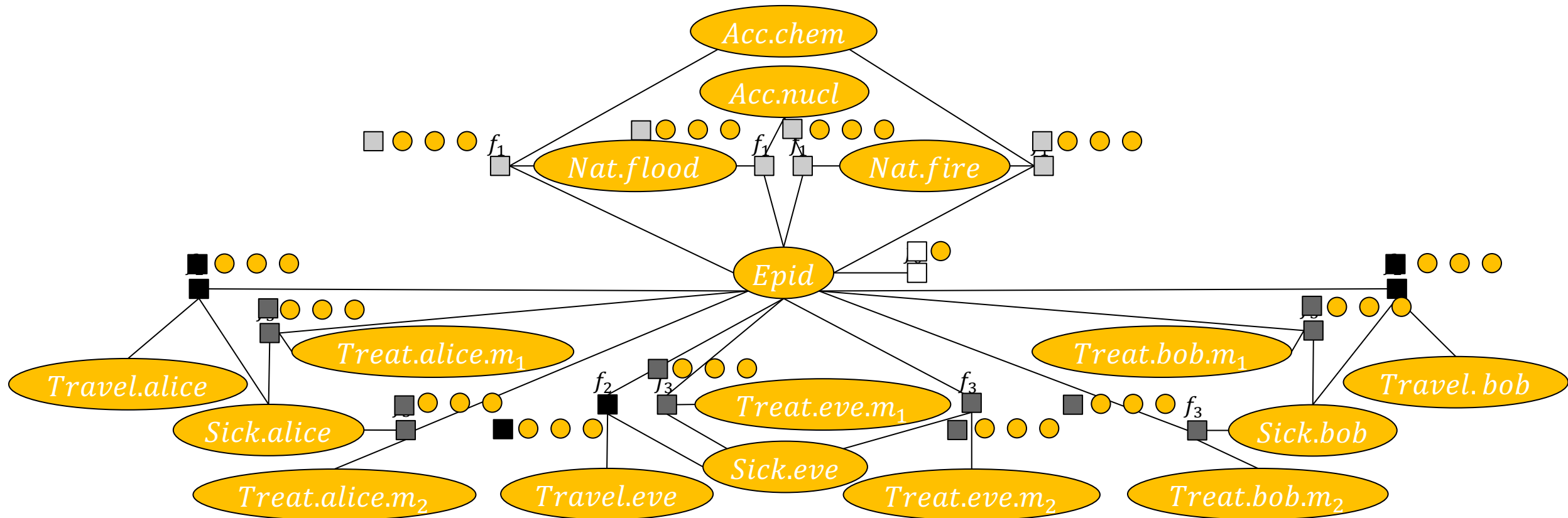
Compression

1. Factors collecting colours from nodes



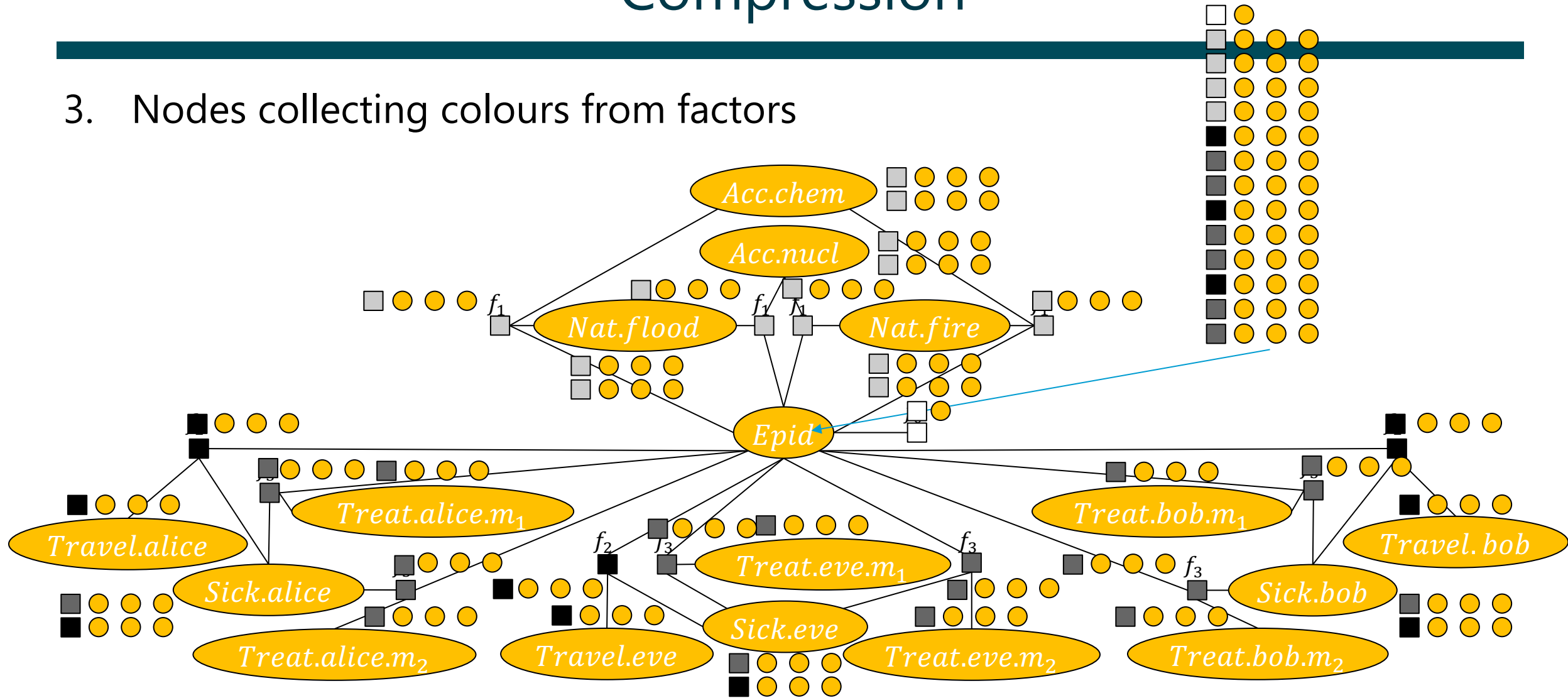
Compression

2. Factors signing their own colours to the collected ones



Compression

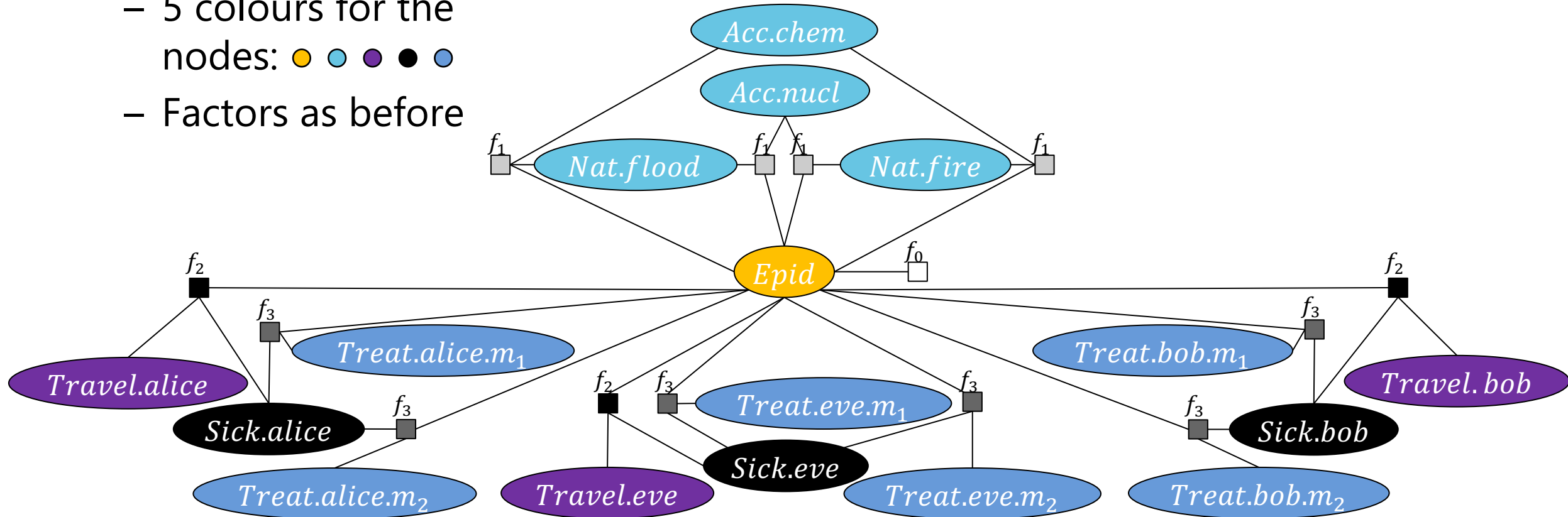
3. Nodes collecting colours from factors



Compression

4. Recolour nodes based on collected signatures

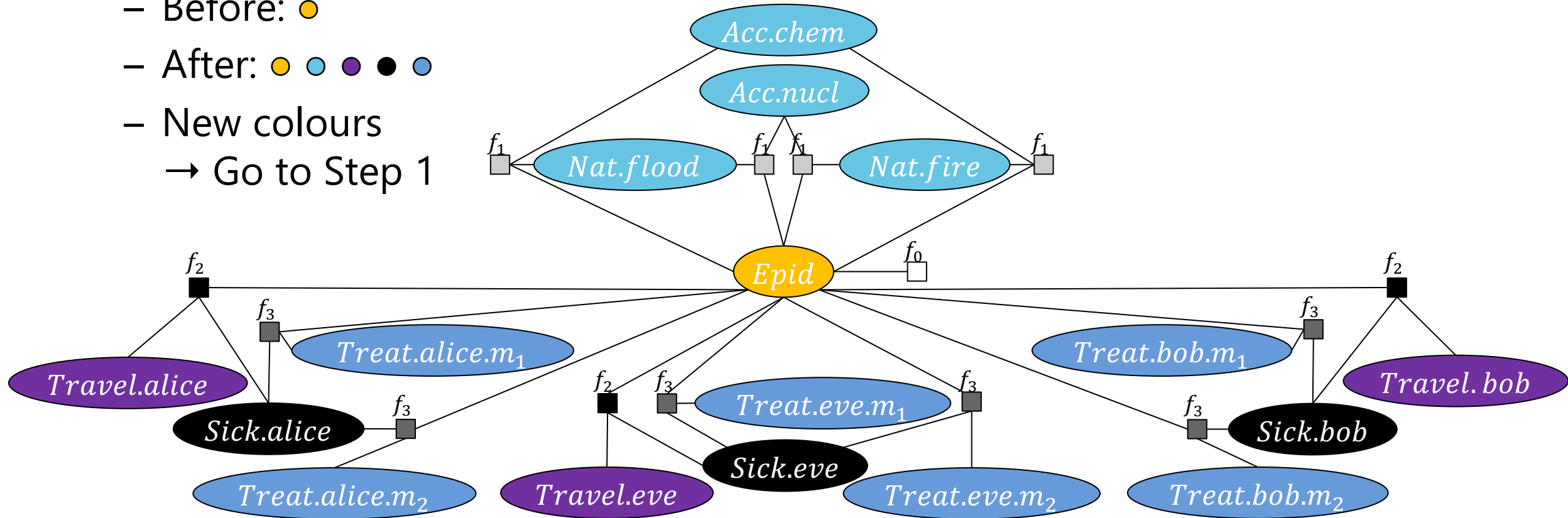
- 5 colours for the nodes: ● ● ● ● ●
- Factors as before



Compression

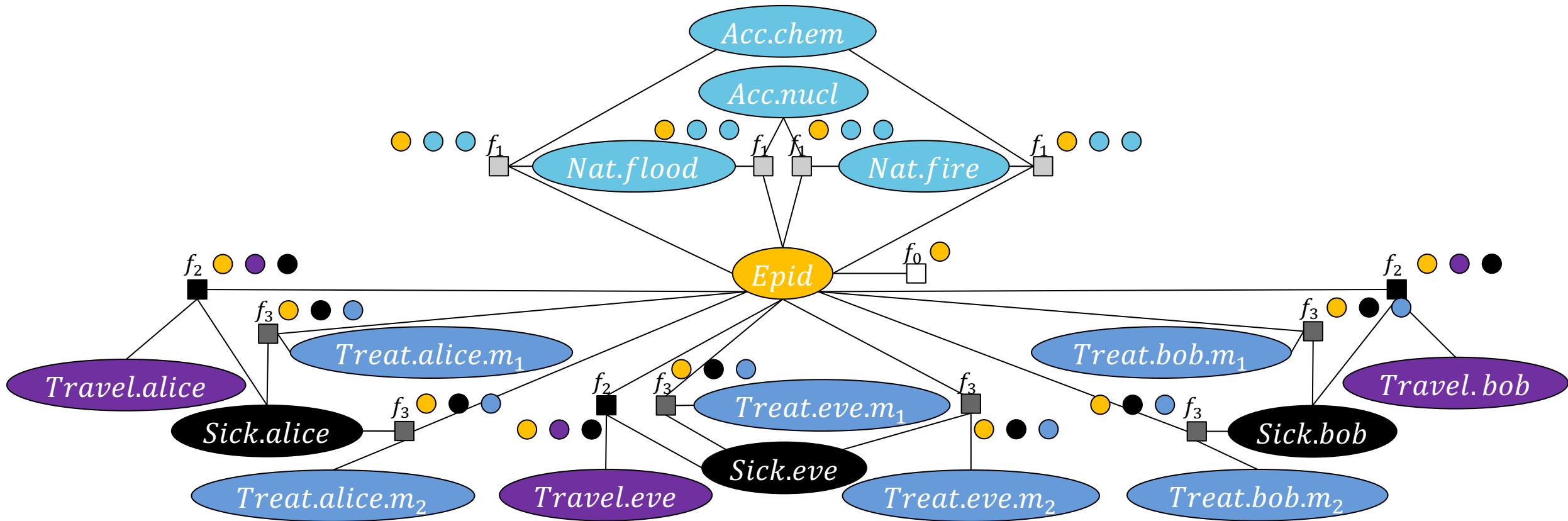
5. If no new colour created, stop. Otherwise, pass colours again.

- Before: ●
- After: ● ● ● ● ●
- New colours
→ Go to Step 1



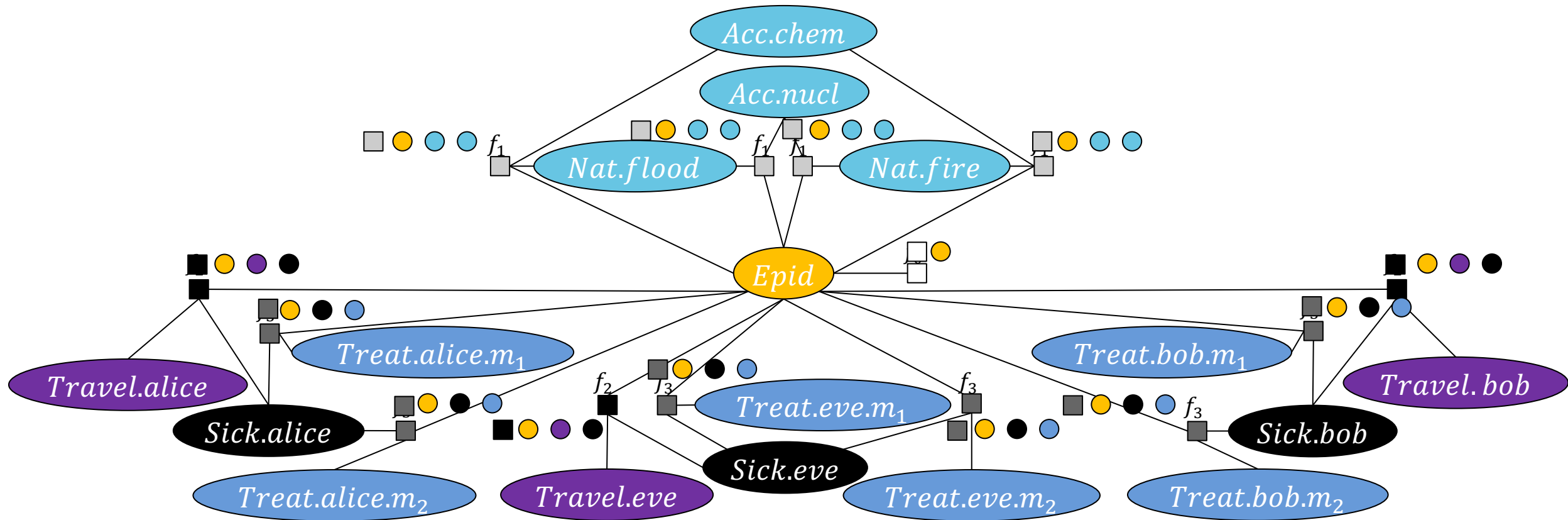
Compression

1. Factors collecting colours from nodes



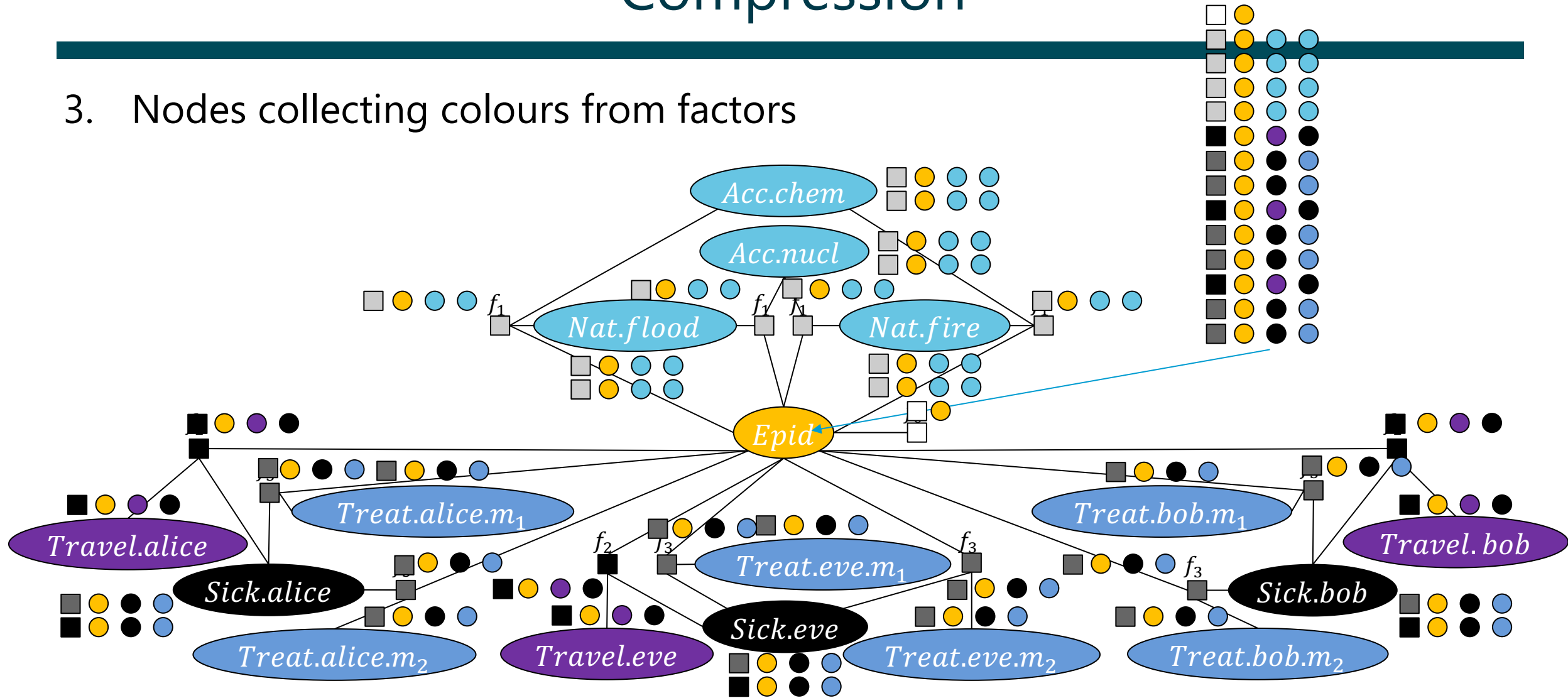
Compression

2. Factors signing their own colours to the collected ones



Compression

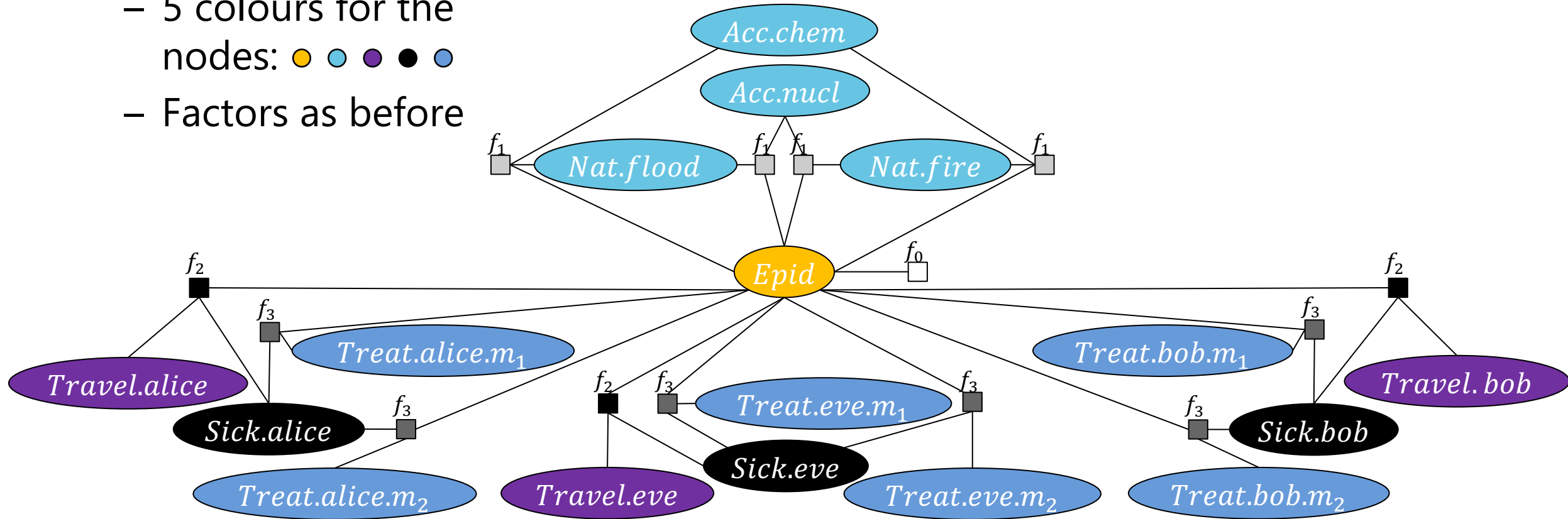
3. Nodes collecting colours from factors



Compression

4. Recolour nodes based on collected signatures

- 5 colours for the nodes: ● ● ● ● ●
- Factors as before



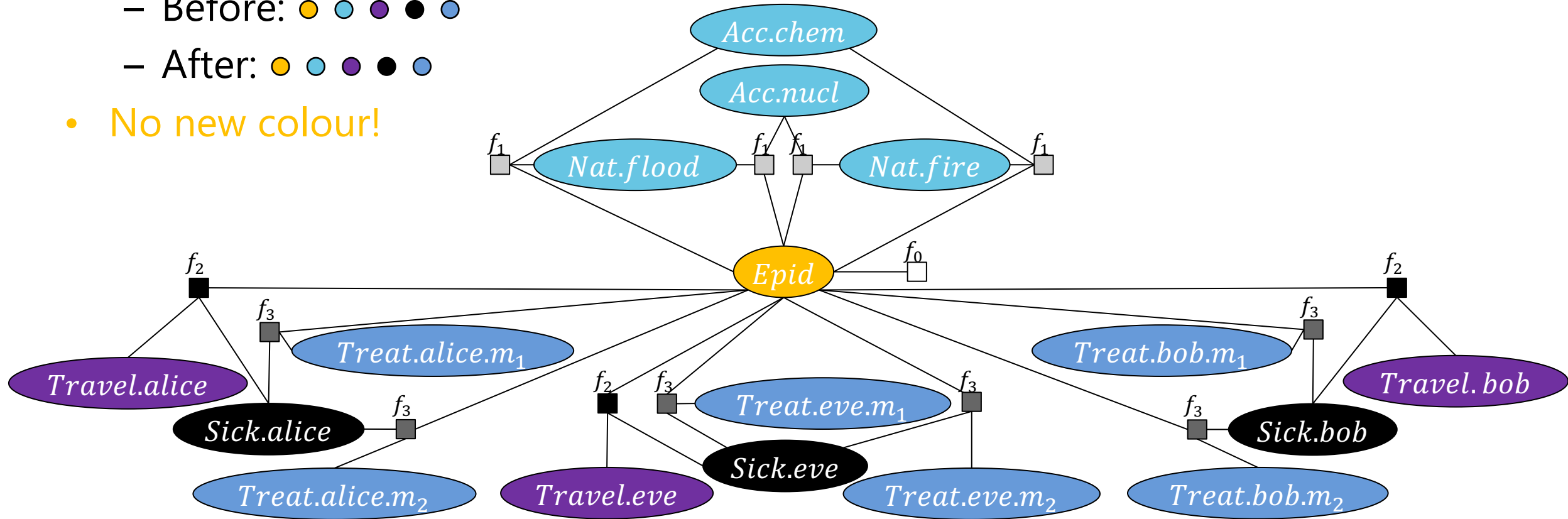
Compression

5. If no new colour created, stop. Otherwise, pass colours again.

– Before: ● ● ● ● ●

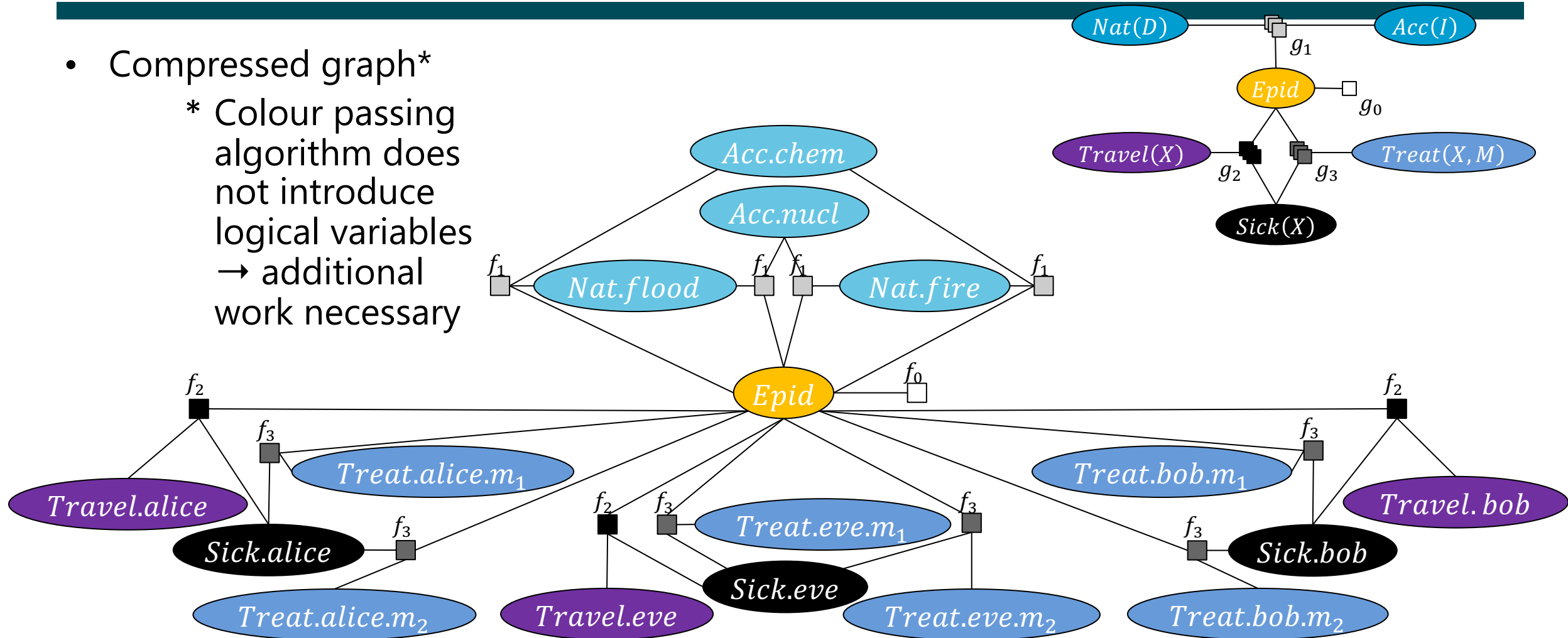
– After: ● ● ● ● ●

• No new colour!



Compression

- Compressed graph*
 - * Colour passing algorithm does not introduce logical variables → additional work necessary



Is the PRM DP?

But does this suffice?

Current state

- Obtain lifted representation
- Groups over indistinguishable objects
- Ground model DP → lifted model DP

Research

- Automatically introduce logical variables
- Use even more symmetries
- Include deniability in colour passing

What about inference?

- Assume DP PRM:
 - $P(S(X))$?
 - $P(S(x_1))$?
 - $P(S(x_1) \mid T(x_1) = \dots)$?
- Exact numbers of individuals per group
- ... time around queries could ... information

Works already well for majority groups!

queries might insert ... the model / ... distinguishable objects ... distinguishable

Evidence Handling

Goal:

- Keep groups (mostly) unchanged
- Keep groups having at least a certain size

Idea:

- Apply (existing) DP Clustering on evidence for different groups
 - On group level suffices
 - Having DP Clustering, we can use the mean to insert clusters as uncertain evidence
 - Impact of one data point limited
 - Minimal group size ensured

Adapting the Query Language

Change:

- Only allow for representative queries
- Perturbate group sizes/construct DP histograms for query answering

Impact:

- $P(S(X))$ returns probability distribution for a representative of each indistinguishable group
- Deniability w.r.t. group sizes, if published

Adding Time

Problem:

- Model can slowly ground
- Individuals should not be used multiple times even with limited data

Idea:

- Use existing methods to approximate symmetries over time to increase group sizes
- Having a representation of individuals, we can rotate cohorts

Generative Model

- Assume DP model and inference
 - PRMs are a generative model
 - Sample from distributions to obtain new data sets
 - Sampled data sets are DP
 - Possibility to share data

Conclusion

- DP PRM
 - Colour passing a good starting point to obtain a lifted DP model
- Out-of-the-box lifting algorithms are a good starting point for DP
 - Groups of indistinguishable objects
 - Representation/ sense of individuals
- Modifications required for sensitive data points
 - Insertion of sensitive data points
 - Hide sensitive information
- Synthesising new data set to be published
 - Using a DP model and DP inference possible